frontiers | Frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 23 May 2022
doi: 10.3389/fbioe.2022.872044

Pretreatment Computed Tomography-Based Machine Learning Models to Predict Outcomes in Hepatocellular Carcinoma Patients who Received Combined Treatment of Trans-Arterial Chemoembolization and Tyrosine Kinase Inhibitor

Qianqian Ren[1,2†], *Peng Zhu*[3†], *Changde Li*[1,2], *Meijun Yan*[1,2], *Song Liu*[1,2], *Chuansheng Zheng*[1,2] *and Xiangwen Xia*[1,2*]

[1]Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Hubei Province Key Laboratory of Molecular Imaging, Wuhan, China, [3]Department of Hepatobiliary Surgery, Wuhan No.1 Hospital, Wuhan, China

OPEN ACCESS

Edited by:
Hung-Yin Lin,
National University of Kaohsiung, Taiwan

Reviewed by:
Angela Lombardi,
Università degli Studi di Bari, Italy
Kranthi Kolli,
Abbott, United States

***Correspondence:**
Xiangwen Xia
xiangwen_xia@hust.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Bionics and Biomimetics, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 09 February 2022
Accepted: 22 April 2022
Published: 23 May 2022

Citation:
Ren Q, Zhu P, Li C, Yan M, Liu S, Zheng C and Xia X (2022) Pretreatment Computed Tomography-Based Machine Learning Models to Predict Outcomes in Hepatocellular Carcinoma Patients who Received Combined Treatment of Trans-Arterial Chemoembolization and Tyrosine Kinase Inhibitor. Front. Bioeng. Biotechnol. 10:872044. doi: 10.3389/fbioe.2022.872044

Aim: Trans-arterial chemoembolization (TACE) in combination with tyrosine kinase inhibitor (TKI) has been evidenced to improve outcomes in a portion of patients with hepatocellular carcinoma (HCC). Developing biomarkers to identify patients who might benefit from the combined treatment is needed. This study aims to investigate the efficacy of radiomics/deep learning features-based models in predicting short-term disease control and overall survival (OS) in HCC patients who received the combined treatment.

Materials and Methods: A total of 103 HCC patients who received the combined treatment from Sep. 2015 to Dec. 2019 were enrolled in the study. We exacted radiomics features and deep learning features of six pre-trained convolutional neural networks (CNNs) from pretreatment computed tomography (CT) images. The robustness of features was evaluated, and those with excellent stability were used to construct predictive models by combining each of the seven feature exactors, 13 feature selection methods and 12 classifiers. The models were evaluated for predicting short-term disease by using the area under the receiver operating characteristics curve (AUC) and relative standard deviation (RSD). The optimal models were further analyzed for predictive performance on overall survival.

Results: A total of the 1,092 models (156 with radiomics features and 936 with deep learning features) were constructed. Radiomics_GINI_Nearest Neighbors (RGNN) and Resnet50_MIM_Nearest Neighbors (RMNN) were identified as optimal models, with the AUC of 0.87 and 0.94, accuracy of 0.89 and 0.92, sensitivity of 0.88 and 0.97, specificity of 0.90 and 0.90, precision of 0.87 and 0.83, F1 score of 0.89 and 0.92, and RSD of 1.30 and 0.26, respectively. Kaplan-Meier survival analysis showed that RGNN and RMNN were associated with better OS (p = 0.006 for RGNN and p = 0.033 for RMNN).

Conclusion: Pretreatment CT-based radiomics/deep learning models could non-invasively and efficiently predict outcomes in HCC patients who received combined therapy of TACE and TKI.

Keywords: radiomics, deep learning, feature robustness, trans-arterial chemoembolization, tyrosine kinase inhibitor, hepatocellular carcinoma

INTRODUCTION

In recent years, many novel therapies have modified the therapeutic landscape of hepatocellular carcinoma (HCC) (A. Rizzo et al., 2021; S. De Lorenzo et al., 2018). Furthermore, predictive biomarkers to guide treatment choice were explored extensively (A. Rizzo and G. Rizzo and Brandi, 2021). In particular, trans-arterial chemoembolization (TACE) combined with tyrosine kinase inhibitor molecular targeted therapy has been shown to significantly improve outcomes over TACE alone in patients with HCC (M. Kudo et al., 2020; Z. Peng et al., 2019). Due to tumor heterogeneity, patients' responses to the combined treatment may vary, indicating exploration of predictors to identify patients who might benefit from the combined treatment is urgently needed (M. Kudo et al., 2020; T. Meyer et al., 2017). Microvascular invasion (MVI) has been proven effective in predicting response to TACE combined with Sorafenib in patients with recurrent intermediate stage HCC (Z. Peng et al., 2019). However, MVI is detected at the resection. Furthermore, tissue-based biomarkers can only reflect the local but not the general characteristics of the heterogeneous nature of the tumor since they mostly rely on a single tumor sample from an approachable lesion in practice. In addition, it is difficult to identify the patient's current status from an archival sample due to the evolution of the tumor and the tumor microenvironment during anti-cancer treatment. The biomarker to identify patients most likely to benefit from this combined treatment is limited.

Radiomics has been used to evaluate the severity of chronic liver disease and assess the prognosis of malignant liver tumors (S. Chen et al., 2019; G. W. Ji et al., 2019; S. Kim et al., 2019; F. Liu et al., 2018; H. J. Park et al., 2019; X. Xu et al., 2019). Deep learning (DL) has been widely applied to liver imaging for various tasks, including organ segmentation, staging liver fibrosis, tumor detection or classification, and improving image quality (C. A. Hamm et al., 2019; F. Liu F et al., 2019; D. Tamada et al., 2020; K. Wang et al., 2019a and X. Liu Z et al., 2019; K. Wang et al., 2019b and A. Mamidipalli et al., 2019; K. Yasaka et al., 2018a, H. Akai, and O. Abe et al., 2018; K. Yasaka et al., 2018b, H. Akai, and A. Kunimatsu et al., 2018). Because training a DL model with a small sample size for one specific clinical question often does not yield satisfactory results, a machine learning framework that combines radiomics features and deep learning features from pre-trained networks with conventional machine learning methods has satisfying predictive performance accuracy and computational costs for some tasks (S. Raghu et al., 2020).

However, the clinical interpretability and reproducibility of clinical-decision support algorithms remain challenging. The robustness of a radiomics/deep-learning-based prediction model refers to its ability to tolerate perturbation to the image input. Recent studies in natural image processing have revealed that the output of DL models can be easily affected by small-scale perturbations added to the input (P. Malhotra et al., 2021; X. Yuan et al., 2019). Correspondingly, many factors are known to induce variability in radiomics features, including noise (D. Mackin et al., 2018), heterogeneous voxel size (M. Shafiq-Ul-Hassan et al., 2018), variability in imaging protocols, different vendors, image reconstruction processes (M. Meyer et al., 2019), Region of Interest (ROI) segmentation (I. Fotina et al., 2012; C. Haarburger et al., 2020; J. Kalpathy-Cramer et al., 2016; Q. Qiu et al., 2019), patient motion, overall image quality as well as tumor phenotype (J. E. van Timmeren et al., 2016).

To the best of our knowledge, this is the first work performing a high-throughput benchmark analysis, along with a feature robustness analysis, to predict short-term tumor response and overall survival in patients with HCC who treated with TACE in combination with targeted molecular therapy.

MATERIALS AND METHODS

Data/Population and Data Acquisition

The ethics committee of our institute approved the study and waived written informed consent due to the retrospective design.

We reviewed the electronic medical records of HCC patients who received combined treatment of TACE and TKI from Sep. 2015 to Dec. 2019 at our institute (Union Hospital, Tongji Medical College, Huazhong University of Science and Technology). The inclusion criteria were as follows: 1) age, ≥ 20 years; 2) tumors confined to the liver without macro-vascular invasion or extra-hepatic metastasis; 3) tumors are measurable by the modified Response Evaluation Criteria in Solid Tumours (mRECIST); 4) Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1, Child-Pugh scores ≤7 points and adequate organ function. Those without complete medical records or high-quality CT images in electronic format were excluded.

Two radiologists reviewed pre-treatment and post-treatment CT images to evaluate short-term tumor response according to the mRECIST. Any inconsistency of assessment results was resolved by consensus. Tumor response was evaluated every 8 weeks. Overall survival (OS) was defined as the time from the date of treatment to the date of death without regarding the cause of death, and censored at the date of last follow-up for survivors. The regimen of TACE plus TKI, response evaluation, clinical data and CT data acquisition are detailed in **Supplementary Material**.

Tumor Segmentation and Imaging Pre-Processing

The Region of interest (ROI) of primary tumor, defined as enhanced area in arterial phase CT images in accordance with mRECIST, was manually delineated by two experienced

radiologists (XW X and QQ R) using a 3D Slicer software (A. Fedorov et al., 2012). To be consistent with deep learning features, three consecutive slices with the maximum cross-sectional area of the tumor lesion were selected. The two observers repeated the same procedures 2 weeks later and any disagreement was resolved through consultation. The brightness, the size and of the image were standardized and the noise in the image was removed using the methods reported in literature (H. Koyuncu and Ceylan, 2018). In brief, resegmentation refers to the process whereby only pixels within a specified grey value range (−1,000, 400) are retained to exclude irrelevant organs and objects. The CT images' appropriate window wide and center were adaptively adjusted based on the tumor region's Hounsfield unit values. The images were then subjected to imaging normalization (the intensity of the image was scaled to 0–255) to avoid data heterogeneity bias. Histogram equalization was used to improve the brightness and contrast of the image for practitioners to analyze. CT images are mainly affected by quantum noise, arising from the variability of the electronic density of tissue voxels, statistically represented by a random Gaussian process. We used Gaussian filter to remove the noise in the image. The images with informative slices (three consecutive axial slices with maximum tumor area) corresponding to the segmented tumor region were cropped to 224 mm × 224 mm using a bounding box spanning the whole tumor area.

Feature Extraction

Six commonly used pre-trained convolutional neural networks (CNNs) (Y. Hu et al., 2021; T. N. Sainath et al., 2015), including InceptionResNetV2, InceptionV3, Resnet50, VGG16, VGG19, and Xception, were pretrained on ImageNet, which contains a large number of object categories and manually annotated training images. When performing deep learning feature extraction, we treated the pre-trained network as an arbitrary feature extractor, allowing the input image to propagate forward, stopping at the pre-specified layer, and taking the outputs of that layer as our features. After removing the last fully connected layer, we got feature maps of CT images with the maximum area of the tumor lesion, which corresponded to location invariance in the input layer. After global pooling, each feature map vector was transformed to a maximal raw value. The representational deep learning features refer to a total of 2048 (Resnet50, InceptionV3, and Xception), 1,536 (InceptionResNetV2) or 512 (VGG16, VGG19) features were converted from feature maps to numeric values.

Handcrafted radiomics features were automatically computed from the radiologist-drawn ROIs using the Pyradiomics package implemented in Python. Defined radiomics features with or without wavelet filtration were extracted in accordance with feature definitions described by the image biomarker standardization initiative (IBSI) reporting guidelines (A. Zwanenburg et al., 2020). Features were divided into three groups: (I) first-order statistics; (II) shape features; and (III) second-order features: gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM), neighborhood gray tone difference matrix (NGTDM).

Feature Robustness Evaluation

The ROI images were adjusted to evaluate the impact of perturbations on feature robustness. We tested three perturbations as follows: 1) slice thickness (S): CT images were reconstructed contiguously at 1, 2, 3 and 5 mm section thicknesses; 2) rotation (R): The image and mask were rotated in the axial (x, y) plane, over a set angle θ [−30°, −15°, 15°, and 30°]; 3) segmentation (Seg): ROIs were automatically expanded or shrinked by 20% (A. Zwanenburg et al., 2019).

The Intra-class Correlation Coefficient ICC was chosen to ensure absolute agreement and not only consistency across perturbations. According to the guidelines (T. K. Koo and Li, 2016), all the features with an ICC of more than 0.85 for all tested perturbations were selected for further study analysis. Raw feature vectors were further standardized by centering on the mean and scaling to unit variance.

Feature Selection of Informative Features and Predictive Model Construction

To further reduce feature dimension, the following steps were performed: 1) removing robust features with zero median absolute deviation (MAD); 2) only considering the top 20% features selected by univariate analysis; 3) algorithm-based feature selection; 4) the wrapper feature selection method based on the recursive feature addition algorithm to select the most predictive features. The features were fed to machine learning classifiers and the performance was evaluated by the area under the receiver operating characteristic curve (AUC). A 10-fold cross validation was used in the feature dimension step to avoid data leakage and overestimation.

The algorithm-based feature selectors included ReliefF (RELF), Fischer Score (FSCR), Gini index (GINI), Chisquare score (CHSQ), joint mutual information (JMI), conditional infomax feature extraction (CIFE), double input symmetric relevance (DISR), mutual information maximization (MIM), conditional mutual information maximization (CMIM), interaction capping (ICAP), *t*-test score (TSCR, only for binary classification), minimum redundancy maximum relevance (MRMR), and mutual information feature selection (MIFS). These selectors take a filter-method approach for feature selection. The filter method filters out the irrelevant feature and redundant columns from the model by using different metrics through ranking.

Twelve supervised machine learning classifiers, including Nearest Neighbors, Support Vector Classifiers (SVC) with linear or radial basis function (RBF) kernels, Gaussian processes, decision trees, random forests, multilayer perceptrons, AdaBoost, naïve Bayes, quadratic discriminant analysis (QDA), XGBoost, and logistic regression, were then used to train models for predicting short-term disease control. These classifiers were all imported from scikit-learn implemented in Python (version 3.6.4) (A. Abraham et al., 2014). During the model debugging, samples were shuffled to ensure data randomization. We adopted the Synthetic minority over-sampling technique (SMOTE) (N. V. Chawla et al., 2002), one

TABLE 1 | Baseline demographic and clinical characteristics of patients.

Characteristic	Total (*n* = 103)	PD (*n* = 31)	Non PD (*n* = 72)	*p* Value
Age (year), (mean ± SD)	52 ± 9	52 ± 8	52 ± 10	0.732
Sex				0.730
Male, n (%)	92 (89.3%)	27 (87.1%)	65 (90.3%)	
Female, n (%)	11 (10.7%)	4 (12.9%)	7 (9.7%)	
ECOG score				0.375
0, n (%)	88 (85.4%)	25 (80.6%)	63 (87.5%)	
1, n (%)	15 (14.6%)	6 (19.4%)	9 (12.5%)	
Aetiology				0.978
Hepatitis B, n (%)	83 (80.6%)	25 (80.6%)	58 (80.6%)	
Hepatitis C, n (%)	14 (13.6%)	4 (12.9%)	10 (13.9%)	
Nonviral hepatitis, n (%)	6 (5.8%)	2 (6.5%)	4 (5.6%)	
Child-Pugh classification				1.000
Child-Pugh A, n (%)	92 (89.3%)	28 (90.3%)	64 (88.9%)	
Child-Pugh B ≤ 7, n (%)	11 (10.7%)	3 (9.7%)	8 (11.1%)	
BCLC stage				0.720
B, n (%)	94 (91.3%)	29 (93.5%)	65 (90.3%)	
C, n (%)	9 (8.7%)	2 (6.5%)	7 (9.7%)	
Maximum tumor diameter(mm), median (range)	59.68 (10.40–153.33)	70.26 (10.40–153.33)	56.10 (13.34–144.21)	0.081
AFP				0.983
≤400 ng/ml, n (%)	53 (51.5%)	16 (51.6%)	37 (51.4%)	
>400 ng/ml, n (%)	50 (48.5%)	15 (48.4%)	35 (48.6%)	

of the commonly-used oversampling algorithms, to achieve class balance during the cross-validation step.

The terminology of each predictive model was consistent with its feature exactor, selector, and classifier. For example, VGG19_FSCR_QDA was a model trained by the QDA classifier, with features selected by FSCR and exacted by VGG19. The predictive performance of the models and their stability was evaluated by the AUC and relative standard deviation (RSD), respectively. RSD was calculated according to the formula: RSD = $(sd_{AUC}/mean_{AUC}) \times 100$, where sd_{AUC} and $mean_{AUC}$ were the standard deviation and mean of the ten cross-validated AUC values, respectively. Accuracy, sensitivity, specificity, precision, and F1 score were also calculated to further evaluate the selected model (Sokolova and Japkowicz, 2006).

Statistical Analysis

Continuous variables with normal distribution were presented as mean ± SD (standard deviation) and those with abnormal distribution were presented as median (range). The continuous variables were compared using the *t* test or Kruskal-Wallis tests. Non-continuous variables were compared using the Pearson X^2 test or Fisher's exact test.

Survival curves were plotted using the Kaplan-Meier method and compared using the log-rank test. Cox proportional hazard analysis was used to identify factors associated with survival. A *p*-value of less than 0.05 was considered statistically significant.

RESULTS

Patient Demographics

A total of 103 HCC patients (92 males and 11 females; age (mean ± SD): 52 ± 9 years) who received combined treatment of TACE and TKI were enrolled in this study. Of these, 72 were identified as disease control (1complete tumor response, 54 partial tumor response, and 17 stable diseases) based on mRECIST, yielding a disease control rate (DCR) of 69.9%. The rest were identified as progressed disease (PD, 30.1%). Clinical and tumor characteristics for all patients are listed in **Table 1**. The clinical and tumor characteristics differences between PD and non-PD groups are statistically insignificant.

Our study setup consists of three parts: 1) feature extraction and robustness analysis; 2) constructing models for predicting disease control, performance analysis, and identification of optimal models; 3) OS prediction performance analysis using the optimal models. **Figure 1** shows the workflow of the study.

Feature Robustness Evaluation

A consistency test was applied to evaluate feature robustness. Imaging perturbations produced a slight impact on the stability of radiomics features, with ICC of 0.93 ± 0.11 for S, 0.94 ± 0.15 for R, and 0.96 ± 0.22 for Seg, respectively. High stability was also observed in S and R perturbations for deep learning features extracted with Rnest50, with ICC of 0.89 ± 0.09 and 0.86 ± 0.12, respectively. However, Seg perturbations had moderate impact on the stability of deep learning features extracted from Rnest50, with an ICC of 0.80 ± 0.14. The results of robustness evaluation for features from all extractors were summarized in **Supplementary Table S1**.

Figure 2 shows the results of feature robustness analysis with ICC cutoff value of 0.85. There were 718/851 (84.37%) robust features in radiomics group. In deep learning group, the highest percentage robust features is 38.87% (199/512) from VGG19 by using the same cutoff value of ICC, followed by 35.11% (719/2048) from Resnet50, 34.38% (176/512) from VGG16, 30.62% (627/2048) from Xception, 13.61% (209/1,536) from InceptionResNetV2, and 7.57% (155/2048) from InceptionV3. The results of features robustness analysis with other ICCs were presented in **Supplementary Figure S1**. These results indicated that radiomics features were more stable than deep

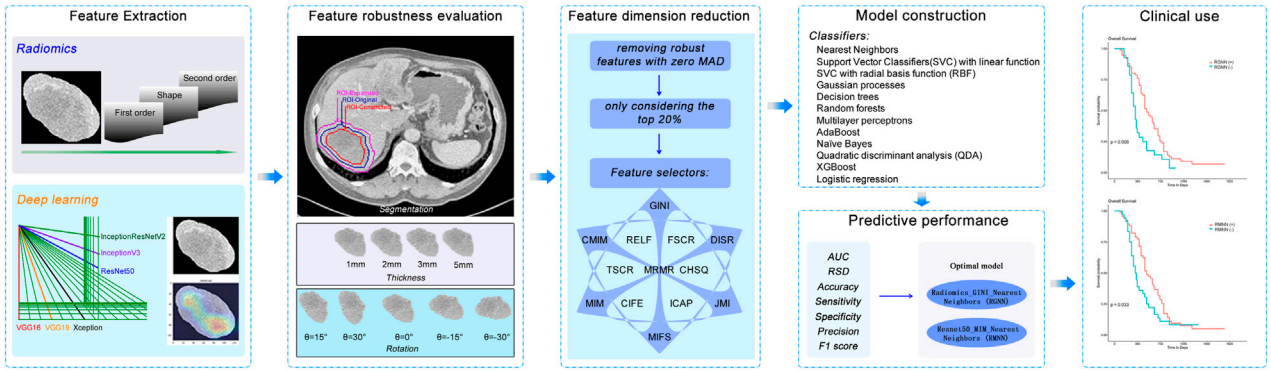


FIGURE 1 | Workflow of major steps in the current work. Tumors are segmented manually and pre-processed. Features are extracted with handcrafted radiomics and six popularly used pre-trained deep learning CNNs, respectively. ICC meters the robustness of features for each perturbation type (segmentation, thickness, and rotation). Robust features are then used to construct models for predicting short-term disease control of tumors by combining each of 13 feature selectors and 12 machine learning classifiers. The best-performing model is evaluated for predicting overall survival.

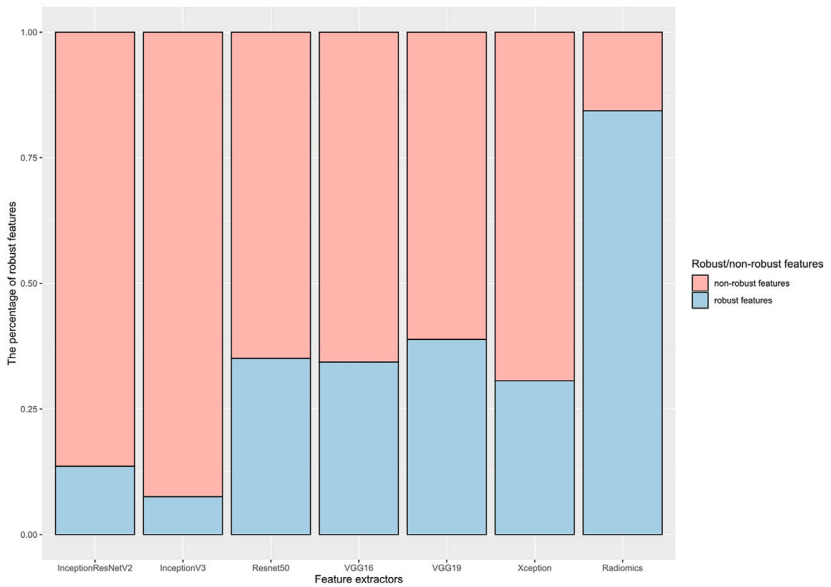


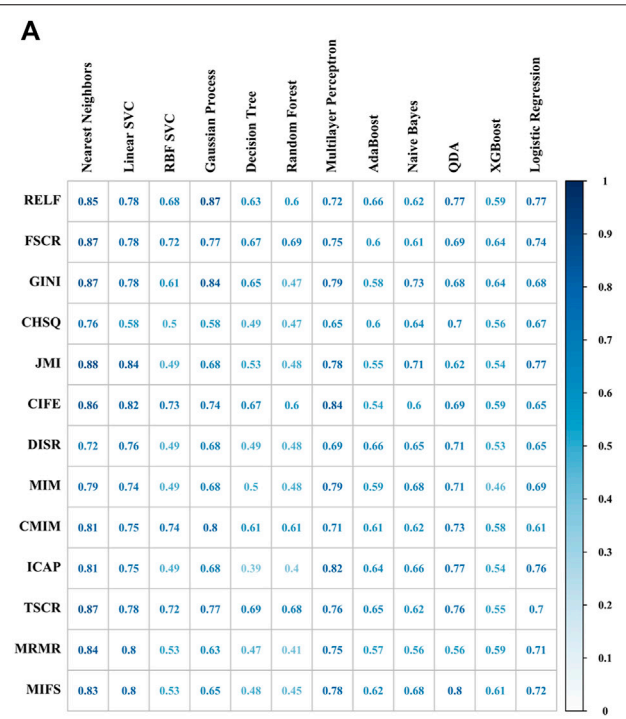
FIGURE 2 | The percentage of robust features against image perturbation.

learning features; in addition, segmentation perturbation (Seg) seemed produce greater impact on stability in deep learning features.

Predictive Performance of Radiomics/Deep Learning Models on Short-Term Disease Control

A total of 156 radiomics features-based models and 936 deep learning features-based models were constructed, and those classified by the k Nearest Neighbors have excellent performance for predicting short-term disease control, reached a median value of AUC of 0.85 (range: 0.64–0.94) (**Supplementary Figure S2**) and median RSD of 1.87 (range: 0.26–11.31).

The Radiomics_GINI_Nearest Neighbors (RGNN) was identified as optimal model in radiomics group, with a cross-validated AUC of 0.87, RSD 1.30, accuracy 0.89, sensitivity 0.88, specificity 0.90, precision 0.87, and F1 score 0.89. (**Figures 3A**,**B**). Radiomics_JMI_Nearest Neighbors had a better AUC value of 0.88, but a higher RSD value of 3.59. The Resnet50_MIM_Nearest Neighbors (RMNN) was identified as the optimal model in deep learning group, with a cross-validated AUC of 0.94, RSD 0.26, accuracy 0.92, sensitivity 0.97, specificity 0.90, precision 0.83, and F1score 0.92 (**Figures 3C,D**). The Resnet50_JMI_Nearest



A

	Nearest Neighbors	Linear SVC	RBF SVC	Gaussian Process	Decision Tree	Random Forest	Multilayer Perceptron	AdaBoost	Naive Bayes	QDA	XGBoost	Logistic Regression
RELF	0.85	0.78	0.68	0.87	0.63	0.6	0.72	0.66	0.62	0.77	0.59	0.77
FSCR	0.87	0.78	0.72	0.77	0.67	0.69	0.75	0.6	0.61	0.69	0.64	0.74
GINI	0.87	0.78	0.61	0.84	0.65	0.47	0.79	0.58	0.73	0.68	0.64	0.68
CHSQ	0.76	0.58	0.5	0.58	0.49	0.47	0.65	0.6	0.64	0.7	0.56	0.67
JMI	0.88	0.84	0.49	0.68	0.53	0.48	0.78	0.55	0.71	0.62	0.54	0.77
CIFE	0.86	0.82	0.73	0.74	0.67	0.6	0.84	0.54	0.6	0.69	0.59	0.65
DISR	0.72	0.76	0.49	0.68	0.49	0.48	0.69	0.66	0.65	0.71	0.53	0.65
MIM	0.79	0.74	0.49	0.68	0.5	0.48	0.79	0.59	0.68	0.71	0.46	0.69
CMIM	0.81	0.75	0.74	0.8	0.61	0.61	0.71	0.61	0.62	0.73	0.58	0.61
ICAP	0.81	0.75	0.49	0.68	0.39	0.4	0.82	0.64	0.66	0.77	0.54	0.76
TSCR	0.87	0.78	0.72	0.77	0.69	0.68	0.76	0.65	0.62	0.76	0.55	0.7
MRMR	0.84	0.8	0.53	0.63	0.47	0.41	0.75	0.57	0.56	0.56	0.59	0.71
MIFS	0.83	0.8	0.53	0.65	0.48	0.45	0.78	0.62	0.68	0.8	0.61	0.72

B

	Nearest Neighbors	Linear SVC	RBF SVC	Gaussian Process	Decision Tree	Random Forest	Multilayer Perceptron	AdaBoost	Naive Bayes	QDA	XGBoost	Logistic Regression
RELF	3.05	2.28	4.38	1.56	7.51	4.43	2.28	8.34	3.88	4.62	4.37	3.41
FSCR	3.34	1.28	5.56	5.54	5.19	3.44	3.31	7.3	4.88	4.48	5.57	4.26
GINI	1.3	1.5	9.5	3.02	11.52	13.63	1.51	6.76	3.03	5.27	6.47	7.59
CHSQ	9.29	9.68	24.45	21.82	10.77	11.03	9.66	15.13	7.5	9.1	9.14	4.92
JMI	3.59	1.16	9.42	6.86	12.35	7.13	2.32	9.35	5.25	2.81	4.75	3.85
CIFE	1.33	1.4	4.11	4.58	10.4	13.38	1.58	7.44	5.73	4.68	11.69	7.77
DISR	6.1	1.41	9.42	6.63	9.45	8.38	3.92	4.89	2.12	3.37	8.43	8.61
MIM	3.34	1.66	9.42	6.86	10.65	6.08	1.31	5.44	4.94	9.6	6.2	1.57
CMIM	1.68	2.49	3.71	2.6	13.32	11.94	2.4	9.84	8.73	3.64	7.03	5.14
ICAP	3.59	1.71	9.42	6.86	5.85	4.62	3.89	9.02	1.85	2.14	4.56	0.79
TSCR	3.34	1.28	5.56	5.54	4.23	4.04	2.27	9.48	13.31	3.39	4.55	5.02
MRMR	3.92	1.31	11.28	5.44	8.55	6.52	3.35	7.09	4.75	8.6	3.53	3.99
MIFS	3.67	1.31	11.28	6.65	12.12	7.25	1.57	10.96	1.75	2.42	12.02	2.47

C

	Nearest Neighbors	Linear SVC	RBF SVC	Gaussian Process	Decision Tree	Random Forest	Multilayer Perceptron	AdaBoost	Naive Bayes	QDA	XGBoost	Logistic Regression
RELF	0.82	0.8	0.81	0.84	0.75	0.73	0.83	0.81	0.84	0.82	0.77	0.82
FSCR	0.9	0.81	0.82	0.92	0.66	0.64	0.86	0.73	0.81	0.91	0.72	0.83
GINI	0.78	0.72	0.66	0.71	0.59	0.57	0.68	0.69	0.85	0.81	0.66	0.73
CHSQ	0.74	0.67	0.66	0.69	0.63	0.64	0.81	0.73	0.79	0.8	0.67	0.82
JMI	0.94	0.86	0.75	0.85	0.8	0.71	0.85	0.87	0.81	0.77	0.82	0.89
CIFE	0.89	0.67	0.74	0.83	0.81	0.77	0.67	0.7	0.8	0.79	0.65	0.72
DISR	0.92	0.81	0.73	0.92	0.82	0.79	0.93	0.72	0.77	0.72	0.73	0.8
MIM	0.94	0.85	0.74	0.85	0.86	0.79	0.88	0.81	0.81	0.81	0.76	0.85
CMIM	0.73	0.58	0.65	0.71	0.61	0.59	0.67	0.61	0.68	0.73	0.57	0.69
ICAP	0.91	0.8	0.8	0.86	0.68	0.67	0.8	0.83	0.79	0.81	0.67	0.79
TSCR	0.9	0.81	0.82	0.92	0.69	0.65	0.8	0.69	0.76	0.74	0.65	0.72
MRMR	0.93	0.82	0.79	0.84	0.8	0.73	0.89	0.88	0.81	0.77	0.76	0.8
MIFS	0.89	0.82	0.79	0.84	0.74	0.7	0.88	0.81	0.69	0.73	0.77	0.78

D

	Nearest Neighbors	Linear SVC	RBF SVC	Gaussian Process	Decision Tree	Random Forest	Multilayer Perceptron	AdaBoost	Naive Bayes	QDA	XGBoost	Logistic Regression
RELF	3.39	3.48	2.01	3.78	6.16	6.03	5.62	8.08	1.64	1.02	4.4	2.4
FSCR	1.55	2.33	1.36	1.18	12.1	7.67	2.45	4.6	3.11	0.5	5.42	2.72
GINI	5.9	6.8	7	9.64	8.5	6.27	6.36	11.38	1.68	4.42	8.12	9.63
CHSQ	2.72	7.12	7.91	7.17	9.94	6.84	1.36	8.82	6.15	4.24	6.09	2
JMI	0.86	1.98	8.5	1.95	4.83	6.05	1.5	2.36	2.59	5.41	4.32	1.5
CIFE	2.64	5.01	5.73	4.23	1.8	5.17	8.35	6.23	6.21	6.6	6.92	6.92
DISR	0.9	1.39	7.82	1.28	5.42	4.84	0.59	10.27	7.54	5.03	4.26	3.25
MIM	0.26	1.69	8.02	1.53	2.55	4.02	2.37	4.11	2.01	1.54	2.53	2.61
CMIM	8.92	16.43	15.97	13.21	15.63	15.69	9.04	9.16	14.43	12.85	13.58	9.14
ICAP	0.69	1.89	7.11	3.24	4.15	6.05	4.62	2.22	4.19	2.16	2.7	2.97
TSCR	1.55	2.33	1.36	1.18	11.2	7.96	2.81	4.33	2.8	2.77	6.29	5.74
MRMR	0.65	2.42	8.05	6.42	3.22	2.94	1.14	2.47	2.1	2.52	1.32	2.33
MIFS	2.6	2.42	8.05	6.42	3.1	3.75	1.37	6.12	5.85	6.29	3.02	6.21

FIGURE 3 | Performance of different combinations of feature selectors (rows) and ML classifiers (columns) for predicting short-term disease control. 10-fold cross-validated AUC values **(A)** and RSD values **(B)** of 156 models with Radiomics features. 10-fold cross-validated AUC values **(C)** and RSD values **(D)** of 156 models with deep learning features extracted from Resnet50.

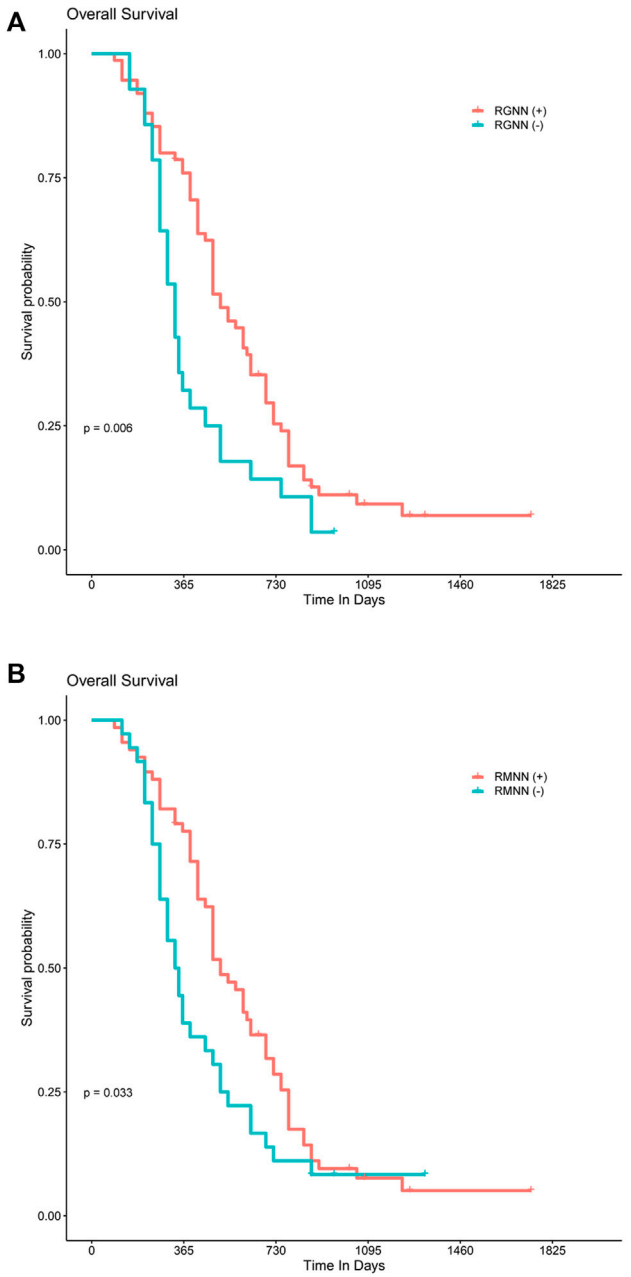


FIGURE 4 | Best-performing model predicting overall survival. Kaplan–Meier survival analysis shows a statistically significant survival advantage for the Radiomics_GINI_Nearest Neighbors **(A)** and Resnet50_MIM_ Nearest Neighbors **(B)**, respectively.

Neighbors had a comparable AUC value 0.94, but a higher RSD value of 0.86.

The list of all feature selectors was in **Supplementary Table S2**; the ML methods' parameter settings and tuning range were presented in **Supplementary Material**. The predictive performance of models constructed by other combinations of CNNs, selectors, and classifiers was **Supplementary Figure S3**.

Predictive Performance of Radiomics_GINI_Nearest Neighbors and Resnet50_MIM_Nearest Neighbors on Overall Survival

For 99 patients with survival data, the median follow-up time was 15 months (range: 10–24 months). The results of the Kaplan-Meier survival analysis are presented in **Figures 4A,B**. There was a statistically significant survival advantage for Radiomics_GINI_Nearest Neighbors (p = 0.006) and Resnet50_MIM_ Nearest Neighbors (p = 0.033). Cox proportional hazard analysis showed that Radiomics_GINI_Nearest Neighbors (HR, 2.49; 95% CI, 1.36–4.55; p = 0.003) and Resnet50_MIM_Nearest Neighbors (HR, 1.83; 95% CI, 1.05–3.17; p = 0.032) was independently associated with overall survival (**Supplementary Table S3**).

DISCUSSION

This study constructed stable radiomics/deep learning models based on a high-throughput analysis for predicting outcomes in HCC patients who received combined treatment of TACE and TKI. We evaluated the robustness of radiomics/deep learning features against multiple perturbations and further evaluated 1,092 combinations of varied feature extractors, selectors, and machine learning techniques. Radiomics_GINI_Nearest Neighbors and Resnet50_MIM_ Nearest Neighbors were identified as the optimal models to predict short-term tumor response and overall survival in two groups (radiomics and deep learning), respectively. Since CT imaging is non-invasive and time-saving, this technique provided us with a fast and auxiliary approach to predict outcomes, thus helping to initially screen patients who might benefit from the combined treatment.

The main idea of deep learning is to employ a deep neural network (DNN) model. To effectively construct deep learning models, we need much more data for training to identify optimal models than prevalent statistical machine learning models. The success of transfer learning schemas, which is frequently used to overcome the limitation of small data sets is clearly contributing to approach DL models as powerful extractors of useful feature sets (H. C. Shin et al., 2016).

Feature robustness depends on the tumor phenotype and is not generalizable (J. E. van Timmeren et al., 2016). In this study, we evaluated the robustness of radiomics and deep learning features by addressing three types of common perturbations, including slice thickness (S), rotation (R), and ROI segmentation (Seg). Our results indicated that Radiomics features seemed more stable than deep learning features in general. To the best of our knowledge, this was the first work to assess the impact of these perturbations on feature stability. The stability of radiomics/deep learning features was more susceptible to Seg. So it is always better performing a "safe" contouring when segmenting, that is, underestimating rather than overestimating the ROI (M. Mottola et al., 2021). These processes can minimize possible variations between centers, machines, image reconstruction methods, and delineation uncertainties. Conducting from these

features, our models can be widely applied for CT data obtained in various institutions.

We further investigated 1,092 combinations of feature exactors, feature selectors, and machine learning techniques to construct predictive models. The DL-based model's prediction ability seemed better than the radiomics-based model. The radiomics-based model with features selected by GINI and classified with Nearest Neighbors was identified to be the optimal model that could effectively predict patients' outcomes. For deep learning-based models, the combination of Resnet50, MIN, and Nearest Neighbors exhibited high predictive power. These results may be helpful for guidance in choosing a better combination of methods. However, which model is better and more practical still needs further studies to verify.

There are several limitations. First, this study was conducted in a single tertiary hospital; limitations inherent to a retrospective design, including small sample size and selection bias, may have influenced the findings. Furthermore, because of the retrospective character of this work, we used perturbation methods rather than test-retest imaging to evaluate feature robustness. In the future, a prospective test-retest study should be conducted. Secondly, there was no external validation cohort to verify the efficacy of our predictive models. Thirdly, three consecutive sections of the tumor were sampled for analysis, and no volume assessment was performed. In a previous study, it was found that data from a single slice was sufficient for this type of analysis (F. Ng et al., 2013). Apart from that, we only investigated some of the influencing factors affecting the image features. Other factors, such as image reconstruction methods, noise removal methods, and histogram equalization approaches, need further studies. Fifthly, although our results demonstrated strong prediction performance, implying that transfer learning might address domain differences, there was heterogeneity across the source and destination databases. Deep learning models explicitly developed for HCC were required. Additionally, the findings' interpretability is a ubiquitous limitation when developing any artificial intelligence or machine learning model applied to medical imaging (F. Cabitza et al., 2017; Z. Liu Z et al., 2019; R. Sun et al., 2018). The issue of findings' interpretability should be improved and solved in further studies.

We believe that the proposed radiomic/deep learning based machine learning model is applicable to other modalities, outcomes, and diseases, with certain modality-specific perturbations. Further research involving standardization across various scanner parameters could aid in harmonizing image attributes in advance. Another major obstacle in this research area is the development of an extensive public database with sufficient annotated medical imaging data to train plenty of parameters in the neural network. Such a database will dramatically help provide more clinically relevant features to train models with better performance.

CONCLUSION

This study constructed stable predictive models from radiomics/deep learning features based on pre-treatment CT imaging using high-throughput analysis. These models could effectively predict short-term tumor response and overall survival in HCC patients who received combined treatment of TACE and targeted molecular therapy. Since CT imaging is non-invasive and time-saving, this technique provided us with a fast and auxiliary approach to identify patients who might benefit from the combined treatment and have the potential to improve precision oncology.

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this article and its online supplementary files. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

QR, PZ, and XX designed the study, wrote the manuscript. CL, MY, SL, and PZ performed retrospective chart reviews. QR, SL, and CZ coordinated Institutional Review Board approval and performed analysis of the data. All authors edited the manuscript. The authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to EditSprings (https://www.editsprings.cn/) for the expert linguistic services provided.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.872044/full#supplementary-material

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine Learning for Neuroimaging with Scikit-Learn. *Front. Neuroinform.* 8, 14. doi:10.3389/fninf.2014.00014

Cabitza, F., Rasoini, R., and Gensini, G. F. (2017). Unintended Consequences of Machine Learning in Medicine. *JAMA* 318 (6), 517–518. doi:10.1001/jama.2017.7797

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). *Smote: Synthetic Minority Over-Sampling Technique. J. Artif. Intell. Res.* 16 (1), 321–357. doi:10.1613/jair.953

Chen, S., Feng, S., Wei, J., Liu, F., Li, B., Li, X., et al. (2019). Pretreatment Prediction of Immunoscore in Hepatocellular Cancer: A Radiomics-Based Clinical Model Based on Gd-Eob-Dtpa-Enhanced Mri Imaging. *Eur. Radiol.* 29 (8), 4177–4187. doi:10.1007/s00330-018-5986-x

De Lorenzo, S., Tovoli, F., Barbera, M. A., Garuti, F., Palloni, A., Frega, G., et al. (2018). Metronomic Capecitabine vs. Best Supportive Care in Child-Pugh B Hepatocellular Carcinoma: A Proof of Concept. *Sci. Rep.* 8 (1), 9997. doi:10.1038/s41598-018-28337-6

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., and Pujol, S. (2012). 3d Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* 30 (9), 1323–1341. doi:10.1016/j.mri.2012.05.001

Fotina, I., Lutgendorf-Caucig, C., Stock, M., Potter, R., and Georg, D. (2012). Critical Discussion of Evaluation Parameters for Inter-Observer Variability in Target Definition for Radiation Therapy. *Strahlenther. Onkol.* 188 (2), 160–167. doi:10.1007/s00066-011-0027-6

Haarburger, C., Muller-Franzes, G., Weninger, L., Kuhl, C., Truhn, D., and Merhof, D. (2020). Radiomics Feature Reproducibility Under Inter-Rater Variability in Segmentations of Ct Images. *Sci. Rep.* 10 (1), 12688. doi:10.1038/s41598-020-69534-6

Hamm, C. A., Wang, C. J., Savic, L. J., Ferrante, M., Schobert, I., Schlachter, T., et al. (2019). Deep Learning for Liver Tumor Diagnosis Part I: Development of a Convolutional Neural Network Classifier for Multi-Phasic Mri. *Eur. Radiol.* 29 (7), 3338–3347. doi:10.1007/s00330-019-06205-9

Hu, Y., Xie, C., Yang, H., Ho, J., Wen, J., Han, L., et al. (2021). Computed Tomography-Based Deep-Learning Prediction of Neoadjuvant Chemoradiotherapy Treatment Response in Esophageal Squamous Cell Carcinoma. *Radiother. Oncol.* 154, 6–13. doi:10.1016/j.radonc.2020.09.014

Ji, G. W., Zhu, F. P., Zhang, Y. D., Liu, X. S., Wu, F. Y., Wang, K., et al. (2019). A Radiomics Approach to Predict Lymph Node Metastasis and Clinical Outcome of Intrahepatic Cholangiocarcinoma. *Eur. Radiol.* 29 (7), 3725–3735. doi:10.1007/s00330-019-06142-7

Kalpathy-Cramer, J., Mamomov, A., Zhao, B., Lu, L., Cherezov, D., Napel, S., et al. (2016). Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. *Tomography* 2 (4), 430–437. doi:10.18383/j.tom.2016.00235

Kim, S., Shin, J., Kim, D. Y., Choi, G. H., Kim, M. J., and Choi, J. Y. (2019). Radiomics on Gadoxetic Acid-Enhanced Magnetic Resonance Imaging for Prediction of Postoperative Early and Late Recurrence of Single Hepatocellular Carcinoma. *Clin. Cancer Res.* 25 (13), 3847–3855. doi:10.1158/1078-0432.CCR-18-2861

Koo, T. K., and Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* 15 (2), 155–163. doi:10.1016/j.jcm.2016.02.012

Koyuncu, H., and Ceylan, R. (2018). Elimination of White Gaussian Noise in Arterial Phase Ct Images to Bring Adrenal Tumours into the Forefront. *Comput. Med. Imaging Graph* 65, 46–57. doi:10.1016/j.compmedimag.2017.05.004

Kudo, M., Ueshima, K., Ikeda, M., Torimura, T., Tanabe, N., Aikata, H., et al. (2020). Randomised, Multicentre Prospective Trial of Transarterial Chemoembolisation (Tace) Plus Sorafenib as Compared with Tace Alone in Patients with Hepatocellular Carcinoma: Tactics Trial. *Gut* 69 (8), 1492–1501. doi:10.1136/gutjnl-2019-318934

Liu, F., Ning, Z., Liu, Y., Liu, D., Tian, J., Luo, H., et al. (2018). Development and Validation of a Radiomics Signature for Clinically Significant Portal Hypertension in Cirrhosis (Chess1701): A Prospective Multicenter Study. *EBioMedicine* 36, 151–158. doi:10.1016/j.ebiom.2018.09.023

Liu, F., Samsonov, A., Chen, L., Kijowski, R., and Feng, L. (2019). Santis: Sampling-Augmented Neural Network with Incoherent Structure for Mr Image Reconstruction. *Magn. Reson. Med.* 82 (5), 1890–1904. doi:10.1002/mrm.27827

Liu, Z., Wang, S., Dong, D., Wei, J., Fang, C., Zhou, X., et al. (2019). The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* 9 (5), 1303–1322. doi:10.7150/thno.30309

Mackin, D., Ger, R., Dodge, C., Fave, X., Chi, P. C., Zhang, L., et al. (2018). Effect of Tube Current on Computed Tomography Radiomic Features. *Sci. Rep.* 8 (1), 2354. doi:10.1038/s41598-018-20713-6

Malhotra, P., Singh, Y., Anand, P., Bangotra, D. K., Singh, P. K., and Hong, W. C. (2021). Internet of Things: Evolution, Concerns and Security Challenges. *Sensors (Basel)* 21 (5), 1809. doi:10.3390/s21051809

Meyer, M., Ronald, J., Vernuccio, F., Nelson, R. C., Ramirez-Giraldo, J. C., Solomon, J., et al. (2019). Reproducibility of Ct Radiomic Features within the Same Patient: Influence of Radiation Dose and Ct Reconstruction Settings. *Radiology* 293 (3), 583–591. doi:10.1148/radiol.2019190928

Meyer, T., Fox, R., Ma, Y. T., Ross, P. J., James, M. W., Sturgess, R., et al. (2017). Sorafenib in Combination with Transarterial Chemoembolisation in Patients with Unresectable Hepatocellular Carcinoma (Tace 2): A Randomised Placebo-Controlled, Double-Blind, Phase 3 Trial. *Lancet Gastroenterol. Hepatol.* 2 (8), 565–575. doi:10.1016/S2468-1253(17)30156-5

Mottola, M., Ursprung, S., Rundo, L., Sanchez, L. E., Klatte, T., Mendichovszky, I., et al. (2021). Reproducibility of Ct-Based Radiomic Features Against Image Resampling and Perturbations for Tumour and Healthy Kidney in Renal Cancer Patients. *Sci. Rep.* 11 (1), 11542. doi:10.1038/s41598-021-90985-y

Ng, F., Kozarski, R., Ganeshan, B., and Goh, V. (2013). Assessment of Tumor Heterogeneity by Ct Texture Analysis: Can the Largest Cross-Sectional Area Be Used as an Alternative to Whole Tumor Analysis? *Eur. J. Radiol.* 82 (2), 342–348. doi:10.1016/j.ejrad.2012.10.023

Park, H. J., Lee, S. S., Park, B., Yun, J., Sung, Y. S., Shim, W. H., et al. (2019). Radiomics Analysis of Gadoxetic Acid-Enhanced Mri for Staging Liver Fibrosis. *Radiology* 290 (2), 380–387. doi:10.1148/radiol.2018181197

Peng, Z., Chen, S., Xiao, H., Wang, Y., Li, J., Mei, J., et al. (2019). Microvascular Invasion as a Predictor of Response to Treatment with Sorafenib and Transarterial Chemoembolization for Recurrent Intermediate-Stage Hepatocellular Carcinoma. *Radiology* 292 (1), 237–247. doi:10.1148/radiol.2019181818

Qiu, Q., Duan, J., Duan, Z., Meng, X., Ma, C., Zhu, J., et al. (2019). Reproducibility and Non-Redundancy of Radiomic Features Extracted from Arterial Phase Ct Scans in Hepatocellular Carcinoma Patients: Impact of Tumor Segmentation Variability. *Quant. Imaging Med. Surg.* 9 (3), 453–464. doi:10.21037/qims.2019.03.02

Raghu, S., Sriraam, N., Temel, Y., Rao, S. V., and Kubben, P. L. (2020). Eeg Based Multi-Class Seizure Type Classification Using Convolutional Neural Network and Transfer Learning. *Neural Netw.* 124, 202–212. doi:10.1016/j.neunet.2020.01.017

Rizzo, A., and Brandi, G. (2021). Biochemical Predictors of Response to Immune Checkpoint Inhibitors in Unresectable Hepatocellular Carcinoma. *Cancer Treat. Res. Commun.* 27, 100328. doi:10.1016/j.ctarc.2021.100328

Rizzo, A., Dadduzio, V., Ricci, A. D., Massari, F., Di Federico, A., Gadaleta-Caldarola, G., et al. (2021). Lenvatinib Plus Pembrolizumab: the Next Frontier for the Treatment of Hepatocellular Carcinoma? *Expert Opin. Investig. Drugs* 31 (4), 371–378. doi:10.1080/13543784.2021.1948532

Sainath, T. N., Kingsbury, B., Saon, G., Soltau, H., Mohamed, A. R., Dahl, G., et al. (2015). Deep Convolutional Neural Networks for Large-Scale Speech Tasks. *Neural Netw.* 64, 39–48. doi:10.1016/j.neunet.2014.08.005

Shafiq-Ul-Hassan, M., Latifi, K., Zhang, G., Ullah, G., Gillies, R., and Moros, E. (2018). Voxel Size and Gray Level Normalization of Ct Radiomic Features in Lung Cancer. *Sci. Rep.* 8 (1), 10545. doi:10.1038/s41598-018-28895-9

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: Cnn Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* 35 (5), 1285–1298. doi:10.1109/TMI.2016.2528162

Sokolova, M., and Japkowicz, N. (2006). Beyond Accuracy, F-Score and Roc: A Family of Discriminant Measures for Performance Evaluation. *Lect. Notes Comput. Sci.* 4304, 1015–1021. doi:10.1007/11941439_114

Sun, R., Limkin, E. J., Vakalopoulou, M., Dercle, L., Champiat, S., Han, S. R., et al. (2018). A Radiomics Approach to Assess Tumour-Infiltrating Cd8 Cells and Response to Anti-pd-1 or Anti-pd-l1 Immunotherapy: An Imaging Biomarker, Retrospective Multicohort Study. *Lancet Oncol.* 19 (9), 1180–1191. doi:10.1016/S1470-2045(18)30413-3

Tamada, D., Kromrey, M. L., Ichikawa, S., Onishi, H., and Motosugi, U. (2020). Motion Artifact Reduction Using a Convolutional Neural Network for Dynamic Contrast Enhanced Mr Imaging of the Liver. *Magn. Reson. Med. Sci.* 19 (1), 64–76. doi:10.2463/mrms.mp.2018-0156

van Timmeren, J. E., Leijenaar, R., van Elmpt, W., Wang, J., Zhang, Z., Dekker, A., et al. (2016). Test-Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific? *Tomography* 2 (4), 361–365. doi:10.18383/j.tom.2016.00208

Wang, K., Lu, X., Zhou, H., Gao, Y., Zheng, J., Tong, M., et al. (2019a). Deep Learning Radiomics of Shear Wave Elastography Significantly Improved Diagnostic Performance for Assessing Liver Fibrosis in Chronic Hepatitis B: A Prospective Multicentre Study. *Gut* 68 (4), 729–741. doi:10.1136/gutjnl-2018-316204

Wang, K., Mamidipalli, A., Retson, T., Bahrami, N., Hasenstab, K., Blansit, K., et al. (2019b). Automated Ct and Mri Liver Segmentation and Biometry Using a Generalized Convolutional Neural Network. *Radiol. Artif. Intell.* 1 (2), 180022. doi:10.1148/ryai.2019180022

Xu, X., Zhang, H. L., Liu, Q. P., Sun, S. W., Zhang, J., Zhu, F. P., et al. (2019). Radiomic Analysis of Contrast-Enhanced Ct Predicts Microvascular Invasion and Outcome in Hepatocellular Carcinoma. *J. Hepatol.* 70 (6), 1133–1144. doi:10.1016/j.jhep.2019.02.023

Yasaka, K., Akai, H., Abe, O., and Kiryu, S. (2018a). Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-Enhanced Ct: A Preliminary Study. *Radiology* 286 (3), 887–896. doi:10.1148/radiol.2017170706

Yasaka, K., Akai, H., Kunimatsu, A., Abe, O., and Kiryu, S. (2018b). Liver Fibrosis: Deep Convolutional Neural Network for Staging by Using Gadoxetic Acid-Enhanced Hepatobiliary Phase Mr Images. *Radiology* 287 (1), 146–155. doi:10.1148/radiol.2017171928

Yuan, X., He, P., Zhu, Q., and Li, X. (2019). Adversarial Examples: Attacks and Defenses for Deep Learning. *IEEE Trans. Neural Netw. Learn Syst.* 30 (9), 2805–2824. doi:10.1109/TNNLS.2018.2886017

Zwanenburg, A., Leger, S., Agolli, L., Pilz, K., Troost, E., Richter, C., et al. (2019). Assessing Robustness of Radiomic Features by Image Perturbation. *Sci. Rep.* 9 (1), 614. doi:10.1038/s41598-018-36938-4

Zwanenburg, A., Vallieres, M., Abdalah, M. A., Aerts, H., Andrearczyk, V., Apte, A., et al. (2020). The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-Based Phenotyping. *Radiology* 295 (2), 328–338. doi:10.1148/radiol.2020191145

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ren, Zhu, Li, Yan, Liu, Zheng and Xia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

GLOSSARY

AUC area under the curve
CHSQ Chisquare score
CIFE conditional infomax feature extraction
CMIM conditional mutual information maximization
DISR double input symmetric relevance
DL deep learning
FSCR Fischer Score
GINI Gini index
GLCM gray level co-occurrence matrix
GLDM gray level dependence matrix
GLRLM gray level run length matrix
GLSZM gray level size zone matrix
HCC hepatocellular carcinoma
IBSI image biomarker standardization initiative
ICAP interaction capping
JMI joint mutual information
MAD median absolute deviation
MIFS mutual information feature selection
MIM mutual information maximization
mRECIST modified Response Evaluation Criteria in Solid Tumours
MRMR minimum redundancy maximum relevance
NGTDM neighborhood gray tone difference matrix
QDA quadratic discriminant analysis
RBF radial basis function
RELF ReliefF
RGNN Radiomics_GINI_Nearest Neighbors
RMNN Resnet50_MIM_Nearest Neighbors
ROI Regions of interest
RSD relative standard deviation in percentile
SVC support Vector Classifiers
TACE Trans-arterial chemoembolization
TKI tyrosinekinase inhibitor
TSCR *t*-test score